



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

<b>(51) Classification internationale des brevets 7 :</b> <b>C12N 15/10, 15/31, 1/21, C07K 14/00</b>	<b>A1</b>	<b>(11) Numéro de publication internationale:</b> <b>WO 00/09679</b> <b>(43) Date de publication internationale:</b> 24 février 2000 (24.02.00)
<b>(21) Numéro de la demande internationale:</b> PCT/FR99/01973 <b>(22) Date de dépôt international:</b> 11 août 1999 (11.08.99) <b>(30) Données relatives à la priorité:</b> 98/10338 12 août 1998 (12.08.98) FR <b>(71) Déposant (pour tous les Etats désignés sauf US):</b> PROTEUS (S.A.) [FR/FR]; 1105, avenue Pierre Mendès France, F-30000 Nîmes (FR). <b>(72) Inventeurs; et</b> <b>(75) Inventeurs/Déposants (US seulement):</b> DUPRET, Daniel [FR/FR]; 8, chemin de Maruejols Sinsans, F-30420 Calvisson (FR). MASSON, Jean-Michel [FR/FR]; 43, chemin Flou de Rious, F-31400 Toulouse (FR). LEFEVRE, Fabrice [FR/FR]; Bâtiment B, appartement 208, 2, rue Théodore Aubanel, F-30900 Nîmes (FR). <b>(74) Mandataires:</b> BREESE, Pierre etc.; Bresse-Majerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).	<b>(81) Etats désignés:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Publiée</b> <i>Avec rapport de recherche internationale.</i>	
<b>(54) Title:</b> METHOD FOR OBTAINING <i>IN VITRO</i> RECOMBINED POLYNUCLEOTIDE SEQUENCES, SEQUENCE BANKS AND RESULTING SEQUENCES <b>(54) Titre:</b> PROCEDE D'OBTENTION <i>IN VITRO</i> DE SEQUENCES POLYNUCLEOTIDIQUES RECOMBINEES, BANQUES DE SEQUENCES ET SEQUENCES AINSI OBTENUES <b>(57) Abstract</b> <p>The invention concerns a method for producing <i>in vitro</i> recombined polynucleotide sequences from a bank of polynucleotide sequences, characterised in that it comprises the following steps: fragmenting an initial bank of double-strand polynucleotide sequences; (b) denaturation of the fragments resulting from step (a) optionally in the presence of one or several assembling matrix or matrices; (c) hybridising said fragments with one or several of the assembling matrix or matrices if the latter is/are not present in step (b); (d) ligation of said fragments; (e) optionally cloning said recombined polynucleotide sequences having advantageous properties with respect to one or several reference sequences.</p> <b>(57) Abrégé</b> <p>La présente invention a pour objet un procédé de production <i>in vitro</i> de séquences polynucléotidiques recombinaées à partir d'une banque de séquences polynucléotidiques, caractérisé en ce qu'il comprend les étapes suivantes: (a) la fragmentation d'une banque initiale de séquences polynucléotidiques double-brins; (b) la dénaturation des fragments issus de l'étape (a) éventuellement en présence d'une ou plusieurs matrice(s) d'assemblage; (c) l'hybridation desdits fragments avec une ou plusieurs matrice(s) d'assemblage si celle(s)-ci n'est (ne sont) pas présente(s) à l'étape (b); (d) la ligation desdits fragments; (e) éventuellement le clonage des séquences polynucléotidiques recombinaées, et la sélection des séquences polynucléotidiques recombinaées, et la sélection des séquences polynucléotidiques présentant des propriétés avantageuses par rapport à une ou plusieurs séquences de référence.</p>		

# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroon	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

PROCÉDÉ D'OBTENTION *IN VITRO* DE SÉQUENCES  
POLYNUCLÉOTIDIQUES RECOMBINÉES, BANQUES DE SÉQUENCES ET  
SÉQUENCES AINSI OBTENUES.

5 La présente invention se rapporte à une méthode  
d'obtention *in vitro* de séquences polynucléotidiques  
recombinées. L'invention vise tout particulièrement à  
générer puis sélectionner des séquences polynucléotidiques  
susceptibles de présenter une ou plusieurs propriétés  
10 avantageuses par rapport aux propriétés correspondantes de  
séquences de référence et donc capables de conférer un  
phénotype amélioré et/ou de produire des protéines  
améliorées.

On entend par séquence de référence une  
15 séquence ayant des propriétés proches de celles  
recherchées.

Différentes techniques ont été développées pour  
favoriser la recombinaison *in vitro* entre différentes  
séquences polynucléotidiques, parmi celles-ci on peut citer  
20 plus particulièrement le DNA-Shuffling (12) et le StEP (14)  
toutes deux fondées sur l'utilisation de la PCR.

Le DNA-Shuffling comporte deux étapes, la  
fragmentation aléatoire par la DNase I de séquences  
polynucléotidiques, et une amplification par PCR dans  
25 laquelle les fragments précédemment générés servent  
d'amorces. A chaque étape d'hybridation, le changement de  
matrice provoque des recombinaisons au niveau des régions  
ayant des séquences homologues. Une représentation  
schématique de cette méthode est donnée à la figure 1A en  
30 annexe.

Le StEP consiste à mélanger différentes  
séquences polynucléotidiques contenant diverses mutations  
en présence d'une paire d'amorces. Ce mélange est soumis à  
une réaction de type PCR dans laquelle les étapes  
35 d'hybridation et de polymérisation sont regroupées en une  
seule étape de très courte durée. Ces conditions permettent  
l'hybridation des amorces mais diminuent la vitesse de  
polymérisation, de façon à ce que les fragments qui sont

partiellement synthétisés s'hybrident aléatoirement sur les séquences polynucléotidiques portant les différentes mutations, permettant ainsi la recombinaison. Une représentation schématique de cette méthode est donnée à la figure 1B en annexe.

Dans chacune de ces deux méthodes, l'étape de polymérisation est indispensable au processus de recombinaison. Ainsi, en fonction des polymérases choisies, cette étape de polymérisation peut être génératrice de mutations supplémentaires non souhaitées. En outre, à partir d'un certain nombre de cycles, le DNA-Shuffling et le STEP reposent sur le principe de l'hybridation d'une "méga-amorce" (6) sur une matrice, ce qui entraîne probablement des difficultés de mise en oeuvre pour des séquences polynucléotidiques dont la taille est supérieure à 1,5 Kpb (11). Enfin, ces deux techniques ne permettent pas de contrôler le taux de recombinaisons, puisque celles-ci se font aléatoirement au cours des étapes successives de polymérisation.

La présente invention vise précisément à palier les inconvénients ci-dessus en offrant une méthode simple de préparation de séquences polynucléotidiques recombinaison, permettant de générer des séquences polynucléotidiques susceptibles de présenter des propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence et donc capables de conférer un phénotype amélioré et/ou de produire des protéines améliorées.

Ce but est atteint grâce à un procédé d'obtention *in vitro* de séquences polynucléotidiques recombinaison à partir d'une banque de séquences polynucléotidiques, aussi désignée ci-après banque initiale, caractérisé en ce qu'il comprend les étapes suivantes :

a) la fragmentation d'une banque de séquences polynucléotidiques double-brins,

b) la dénaturation des fragments éventuellement en présence d'une ou plusieurs matrice(s) d'assemblage,

c) l'hybridation desdits fragments avec une ou plusieurs matrice(s) d'assemblage si celle(s)-ci n'est (ne sont) pas présente(s) à l'étape (b),

d) la ligation desdits fragments pour obtenir des séquences polynucléotidiques recombinaées,

e) la sélection des séquences polynucléotidiques recombinaées présentant des propriétés avantageuses par rapport aux propriétés correspondantes à une ou plusieurs séquences de référence.

Le procédé de l'invention peut comprendre à l'issue de l'étape (d) et avant l'étape (e), la répétition des étapes (b), (c) et (d) non plus avec les fragments de l'étape (a) mais avec les produits de l'étape (d).

Ce mode de réalisation est particulièrement utile dans le cas où à l'issue de l'étape (d) tous les fragments ne sont pas ligüés. Dans ce cas, le procédé de l'invention comprend en outre, à la fin de l'étape (d) et avant l'étape (e), un ou plusieurs cycles des réactions suivantes:

- dénaturation des fragments ligüés et non ligüés issus de l'étape (d), éventuellement en présence d'une ou plusieurs matrice(s) d'assemblage,

- hybridation desdits fragments avec une ou plusieurs matrice(s) d'assemblage si celle(s)-ci n'est (ne sont) pas présente(s) lors de la dénaturation,

- ligation desdits fragments.

Ces réactions de dénaturation, hybridation et ligation, sont équivalentes aux étapes (b), (c) et (d), mais réalisées non pas avec les fragments de l'étape (a) mais avec les fragments ligüés et non ligüés issus de l'étape (d).

Le procédé de l'invention peut comprendre en outre, une ou plusieurs des étapes suivantes :

- la séparation des séquences polynucléotidiques recombinaées de la ou des matrice(s) d'assemblage avant l'étape (e),

- l'amplification des séquences polynucléotidiques recombinaées double brin avant l'étape (e).

5 - le clonage de séquences polynucléotidiques recombinaées éventuellement après séparation des brins recombinaés de la ou des matrices et obtention du double brin correspondant avant l'étape (e).

10 Les extrémités des fragments générés à l'étape (a) sont telles qu'il puisse y avoir hybridation adjacente de ces extrémités sur la ou les matrice(s) d'assemblage à l'étape (c) et ligation de ces fragments les uns avec les autres à l'étape (d). Les séquences polynucléotidiques de la banque sur laquelle est effectuée le procédé de  
15 l'invention doivent présenter des zones d'homologie soit entre elles, soit avec les matrices d'assemblage, permettant de générer des extrémités de fragments telles que décrites ci-dessus.

20 Une forme de mise en oeuvre avantageuse du procédé de l'invention consiste à réaliser simultanément les étapes (c) et (d) selon une réaction dite de RLR pour l'expression anglaise de "Recombining Ligation Reaction".

25 Outre les avantages indiqués précédemment, le procédé de l'invention est remarquable en ce qu'il favorise et accélère la recombinaison aléatoire *in vitro* de séquences polynucléotidiques, ces séquences polynucléotidiques pouvant être des gènes. On entend par  
30 gène un fragment ou une séquence d'ADN associée à une fonction biologique. Un gène peut être obtenu de différentes façon, dont la synthèse chimique, la synthèse par polymérisation ou par extraction dudit gène à partir d'une source d'acides nucléiques.

35 La recombinaison *in vitro* des séquences polynucléotidiques de la banque initiale par le procédé de l'invention permet donc d'obtenir une nouvelle banque

contenant des séquences ayant acquis une ou plusieurs caractéristiques des séquences de la banque précédente. Le procédé de l'invention constitue donc une technique d'évolution *in vitro*.

5 Le procédé de l'invention constitue une alternative à la PCR recombinatoire telle que mise en oeuvre dans les techniques de DNA shuffling (12) ou de STEP (14), puisqu'il ne nécessite pas d'étape de polymérisation  
10 *in vitro* pour aboutir à la recombinaison. Au contraire, l'étape clef du procédé de l'invention est l'étape (d) de ligation sur une matrice d'assemblage, ce qui assure un très haut degré de fidélité au cours des événements de recombinaison.

15 Le procédé de l'invention est remarquable en ce qu'il permet d'augmenter considérablement l'efficacité du réassemblage des fragments à liguer. En effet, dans le cas d'une séquence découpée en  $n$  fragments, il existe  $n^n$  possibilités de réassociation des fragments en utilisant un  
20 procédé de ligation classique (sans utilisation d'une matrice de réassemblage qui oriente la ligation), parmi lesquelles une seule forme est intéressante. Dans le cas du procédé de l'invention, la ligation est orientée par la matrice d'assemblage, ce qui permet d'obtenir directement  
25 la seule forme intéressante.

La fragmentation de ces séquences polynucléotidiques à l'étape (a) peut se faire soit de manière contrôlée, soit de manière aléatoire.

30 Dans le cas d'une fragmentation réalisée de manière contrôlée, la fragmentation permet de maîtriser avec précision le degré de recombinaison voulu et la position des points de recombinaison. Selon une forme de réalisation préférée du procédé de l'invention, l'étape (a)  
35 consiste à soumettre les séquences polynucléotidiques de la banque à une hydrolyse par l'action d'une ou plusieurs enzymes de restriction. Ainsi, dans une forme de mise en œuvre particulière du procédé de l'invention, le degré de

recombinaison et la position des points de recombinaison des séquences polynucléotidiques recombinaison sont déterminés par la fragmentation de l'étape (a).

5                   Ainsi, plus le nombre de fragments générés par séquence est grand, plus le nombre de fragments nécessaires pour recomposer une séquence est élevé, ce qui entraîne un taux de recombinaison élevé. En outre, la nature et la position des extrémités des fragments générés dans ce mode de réalisation du procédé de l'invention peuvent être  
10 connues et contrôlées, ce qui permet :

- de contrôler avec précision les zones où la recombinaison a lieu, ou

- d'induire la recombinaison entre des  
15 séquences polynucléotidiques, par exemple des gènes, si les extrémités des fragments sont créées dans des zones d'homologie entre ces séquences, ou des zones d'homologies entre ces séquences et la ou les matrices d'assemblage.

20                   Dans le cas d'une fragmentation aléatoire, tout moyen enzymatique ou mécanique connu de l'homme de métier capable de couper aléatoirement l'ADN double brin pourra être utilisé, comme par exemple une digestion par la Dnase I ou une sonication.

25                   Le procédé de l'invention permettant d'augmenter considérablement l'efficacité de réassemblage des fragments à liguer, il peut donc être appliqué à l'orientation de la ligation multi-moléculaire à bouts  
30 francs. Dans cette application, on utilise comme matrice d'assemblage aux étapes (b) ou (c) des oligonucléotides simple ou double brin juste complémentaires de l'extrémité 3' d'un fragment et 5' du fragment adjacent, ce qui permet l'hybridation adjacente de ces deux extrémités sur la même  
35 matrice après l'étape de dénaturation. Une fois hybridées, les extrémités des fragments peuvent être liguées entre-elles de façon à orienter le sens de ligation des fragments



à bout francs. La même approche peut être envisagée pour l'orientation de la ligation de fragments à bouts cohésifs.

Une mise en œuvre toute préférée du procédé de l'invention consiste à ajouter à l'étape (c) et/ou à l'étape (d) des enzymes capables de reconnaître et de couper de manière spécifique les extrémités non hybridées de fragments, lorsque celles-ci recouvrent d'autres fragments hybridés sur la même matrice. Un exemple préféré de ce type d'enzyme est l'enzyme Flap endonucléase (10).

Une mise en œuvre particulière du procédé de l'invention consiste donc à utiliser des enzymes du type Flap endonucléases lorsque les fragments générés à l'étape (a) peuvent se recouvrir lors de l'hybridation sur la matrice d'assemblage à l'étape (c).

Ainsi, lors de l'hybridation de fragments d'ADN simple brin sur une matrice, ces enzymes ont pour propriété de reconnaître et de couper de manière spécifique les extrémités non hybridées de ces fragments, lorsque celles-ci recouvrent d'autres fragments hybridés sur la même matrice. Au cours de l'étape (c) d'hybridation, ces enzymes permettent donc d'augmenter le nombre d'extrémités adjacentes pouvant être liguées à l'étape (d), ce qui est particulièrement important dans le cas de fragments obtenus par coupure aléatoire, car ces fragments présentent des zones de recouvrement les uns avec les autres lorsqu'ils s'hybrident sur la matrice d'assemblage.

Dans une mise en œuvre particulière du procédé de l'invention utilisant une ligase active à haute température et préférentiellement thermostable à l'étape (d), les endonucléases capables de reconnaître et de couper de manière spécifique les extrémités non hybridées des fragments comme les Flaps, ajoutées à l'étape (c) et/ou à l'étape (d) auront les mêmes propriétés de thermorésistance et d'activité à haute température que ladite ligase.

La banque de séquences polynucléotidiques sur laquelle est effectuée le procédé de l'invention peut être générée par toute méthode connue de l'homme du métier, par exemple à partir d'un gène sauvage par étapes de mutagenèse dirigée successives, par PCR "error prone" (2), par mutagenèse aléatoire chimique, par mutagenèse aléatoire *in vivo*, ou en combinant des gènes de familles proches ou distinctes au sein de la même espèce ou d'espèces différentes de façon à disposer d'une variété de séquences polynucléotidiques dans ladite banque.

Parmi ces techniques, l'invention envisage plus particulièrement, un procédé dans lequel la banque de séquences polynucléotidiques double-brins est obtenue par une réaction de polymérisation en chaîne réalisée dans des conditions qui permettent de créer des mutations ponctuelles aléatoires.

La banque initiale de séquences polynucléotidiques double-brins peut être constituée de séquences synthétiques qui seront fragmentées à l'étape (a) ou qui peuvent constituer les fragments de l'étape (a)

Selon une forme de réalisation préférée du procédé de l'invention, l'étape (a) consiste à soumettre les séquences polynucléotidiques de la banque à une hydrolyse par l'action d'une ou plusieurs enzymes de restriction.

Pour accroître le degré de recombinaison généré par le procédé de l'invention, il suffit d'augmenter le nombre de fragments de restriction en utilisant des enzymes de restriction ayant un grand nombre de sites de coupure sur les séquences polynucléotidiques de la banque, ou en combinant plusieurs enzymes de restriction. Dans le cas de l'utilisation d'une ligase thermostable et thermoactive, la taille du plus petit fragment ainsi généré sera avantageusement supérieure ou égale à 40 pb, afin de conserver une température d'hybridation compatible avec

celle de l'étape (d) de ligation qui est généralement de l'ordre de 65 °C.

5 L'étape (a) peut encore être réalisée en  
généralant une banque de fragments par traitement aléatoire  
enzymatique ou mécanique. En particulier, l'étape (a) peut  
consister en un traitement aléatoire avec la DNase I d'une  
banque de séquences polynucléotidiques double brins  
partiellement hétérologues. Dans le cas où l'on utilise à  
10 l'étape (a) une fragmentation enzymatique ou mécanique  
aléatoire, cette forme de mise en oeuvre du procédé de  
l'invention a la particularité de permettre l'utilisation  
des fragments générés par ce traitement comme matrices les  
uns pour les autres, pour l'hybridation au cours de  
15 l'étape (c) ou de la réaction de RLR des étapes (c) et (d)  
simultanées.

L'étape (b) peut être réalisée en combinant au  
moins deux banques de fragments distinctes générées  
20 séparément à l'étape (a) à partir de la même banque  
initiale par des traitements différents, comme par exemple  
avec des enzymes de restriction différentes. Dans le cas de  
la mise en oeuvre de telles banques, on utilise les  
fragments obtenus à l'étape (a) comme matrices les uns pour  
25 les autres, pour l'hybridation au cours de l'étape (c) ou  
de la réaction de RLR des étapes (c) et (d) simultanées.

Les fragments de l'étape (a) du procédé de  
l'invention peuvent également être générés par des  
30 réactions d'amplification comme la PCR menées sur les  
séquences polynucléotidiques de la banque. Deux solutions  
sont envisageables. Dans un premier cas, les  
oligonucléotides amorces peuvent être conçus de manière à  
générer des fragments dont les extrémités sont adjacentes  
tout au long de la séquence d'assemblage. Dans un deuxième  
35 cas, les oligonucléotides amorces sont conçus de façon à  
générer des fragments ayant des séquences communes, ces  
fragments pouvant servir de matrice d'assemblage les uns

pour les autres à l'étape (b) ou à l'étape (c). Le procédé de l'invention permet de combiner de manière aléatoire différents fragments obtenus à l'étape (a) et de les réassembler au cours des étapes (b), (c) et (d) au sein d'une séquence polynucléotidique. Ce procédé reproduit donc *in vitro* les phénomènes de recombinaison qui peuvent se produire *in vivo* en les favorisant. Le procédé de l'invention est donc tout particulièrement intéressant pour recombinaison des séquences polynucléotidiques entre elles afin de générer de nouvelles séquences polynucléotidiques ayant des propriétés intéressantes par rapport aux propriétés correspondantes de séquences de référence.

L'efficacité de recombinaison du procédé de l'invention est fonction du nombre de fragments générés par séquence polynucléotidique à l'étape (a). En conséquence, le procédé de l'invention utilisera des séquences polynucléotidiques ayant été fragmentés en *n* fragments, *n* étant avantageusement supérieur ou égal à trois.

La matrice d'assemblage à l'étape (b) ou (c) est par exemple une séquence polynucléotidique issue de la banque initiale ou une séquence consensus de ladite banque, simple ou double brin. Dans le cas où la matrice d'assemblage est incorporée directement à l'étape (c) de l'invention, cette matrice doit être sous forme simple brin.

Selon une variante du procédé de l'invention, les matrices d'assemblage des étapes (b) ou (c) sont constituées d'oligonucléotides simples ou doubles brins.

Selon une forme particulière de mise en oeuvre du procédé de l'invention, des oligonucléotides, simple ou double brin, de longueur variable, sont ajoutés à l'étape (b) ou (c) en plus de la matrice. Ces oligonucléotides sont conçus pour pouvoir se substituer à une partie des fragments à l'étape (c), en effet, leur séquence est telle que :

- s'ils sont parfaitement homologues avec la séquence du fragment qu'ils remplacent, ils favorisent certaines combinaisons, ou

5       - s'ils sont partiellement hétérologues avec la séquence du fragment qu'ils remplacent, ils introduisent une ou des mutations dirigées supplémentaires.

10       Avant l'étape (e) du procédé de l'invention, il est possible de séparer les séquences polynucléotidiques recombinaées de la matrice d'assemblage grâce à un marqueur présent sur la matrice d'assemblage ou sur les séquences polynucléotidiques recombinaées. Il est en effet possible de marquer chaque brin de la matrice selon des techniques connues de l'homme du métier. Par exemple, le marqueur de  
15       la matrice d'assemblage peut être un haptène et l'on sépare les séquences polynucléotidiques recombinaées de la matrice d'assemblage par des techniques connues de l'homme du métier, comme par exemple un anticorps anti-haptène fixé sur un support ou une réaction biotine-streptavidine, si  
20       l'haptène est un marqueur biotine.

25       D'autres techniques peuvent être employées pour séparer les séquences polynucléotidiques recombinaées de la matrice d'assemblage. La matrice d'assemblage peut aussi être préparée spécifiquement de façon à faciliter son élimination à la fin du procédé de l'invention. Elle peut ainsi être synthétisée par amplification PCR utilisant des dATP méthylés, ce qui permet sa dégradation par l'endonucléase de restriction Dpn I. Dans ce cas, les séquences polynucléotidiques recombinaées ne doivent pas  
30       contenir de dATP méthylés. La matrice peut aussi avoir été préparée par amplification PCR en utilisant des dUTP, ce qui permet sa dégradation par traitement avec une uracyl-DNA-glycosylase. A l'inverse, il est possible de protéger les séquences polynucléotidiques recombinaées en les  
35       amplifiant par PCR sélective avec des oligonucléotides portant des groupements phosphorothioates en 5'. Un traitement avec une exonucléase permet alors de dégrader spécifiquement la matrice d'assemblage.

Le procédé de l'invention peut comprendre avant le clonage éventuel de l'étape (e), une étape d'amplification des séquences polynucléotidiques recombinaées. Toute technique d'amplification est acceptable notamment une amplification par PCR. Une des plus simple consiste à réaliser une PCR qui permet d'amplifier spécifiquement les séquences polynucléotidiques recombinaées grâce à des amorces qui ne peuvent s'hybrider que sur les extrémités des séquences recombinaées. Les produits PCR sont ensuite clonés pour être caractérisés et les séquences polynucléotidiques présentant des propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence sont sélectionnées.

L'invention a pour objet de générer des séquences polynucléotidiques susceptibles de présenter des propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence. Les séquences polynucléotides recombinaées obtenues à l'étape (d) et éventuellement clonées sont criblées par tout moyen approprié pour sélectionner les séquences polynucléotidiques recombinaées ou les clones présentant des propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence. On entend, par exemple, par propriétés avantageuses la thermostabilité d'une enzyme ou sa qualité à pouvoir fonctionner dans des conditions de pH ou de température ou de concentration saline plus adaptées à un procédé enzymatique, que les protéines témoins habituellement utilisées pour ledit procédé. A titre d'exemple d'un tel procédé, on peut citer un procédé industriel de désencollage des fibres textiles ou de blanchiment des pâtes à papier ou de la production d'arômes dans l'industrie laitière, les procédés de biocatalyse pour la synthèse par voie enzymatique de nouvelles molécules thérapeutiques, etc..

Selon un mode de réalisation avantageux du procédé de l'invention, la banque de séquence

polynucléotidique peut donc être le résultat d'un crible ayant permis de sélectionner par tout moyen approprié les séquences polynucléotidiques présentant des propriétés avantageuses par rapport à des séquences témoins. Les séquences ainsi sélectionnées constituent une banque restreinte.

Mais, il est aussi possible de partir d'une banque non restreinte afin de conserver la représentativité des propriétés contenues dans cette banque.

Les séquences codant pour la ou les protéines présentant une ou des propriétés avantageuses par rapport aux protéines de référence sont alors sélectionnées, par des criblages *in vivo* ou *in vitro*, et peuvent servir à constituer une nouvelle banque pour une éventuelle réitération du procédé de l'invention. Une mise en œuvre avantageuse du procédé de l'invention consiste donc à utiliser comme banque plusieurs séquences polynucléotidiques sélectionnées après une première mise en œuvre du procédé de l'invention, éventuellement mélangées avec d'autres séquences polynucléotidiques. Parmi les techniques de criblage qui peuvent être appliquées à chacun des clones de l'étape (e), les techniques de criblage *in vitro* présentent l'avantage de s'affranchir des problèmes de physiologie cellulaire, et de tous les inconvénients liés au clonage d'expression *in vivo*. En outre, ce type de criblage est facilement automatisable, ce qui permet de cribler un nombre élevé de séquences polynucléotidiques recombinaées.

L'invention concerne aussi une séquence polynucléotidique recombinaée obtenue par un procédé selon l'invention, ainsi qu'un vecteur contenant une telle séquence polynucléotidique recombinaée, un hôte cellulaire transformé par une séquence polynucléotidique recombinaée ou un vecteur de l'invention, ainsi qu'une protéine codée par cette séquence polynucléotidique recombinaée. L'invention comprend également les banques correspondantes de séquences

polynucléotidiques recombinaées, de vecteurs, d'hôtes cellulaires ou de protéines.

5 D'autres avantages et caractéristiques de l'invention apparaîtront des exemples de réalisation de l'invention qui suivent et qui se réfèrent aux dessins en annexe dans lesquels :

10 La figure 1 est une représentation schématique des procédés de l'art antérieur correspondant respectivement au DNA-shuffling (figure 1A) et au StEP figure 1B).

La figure 2 est une représentation schématique d'un exemple de réalisation du procédé de l'invention et de certaines de ses variantes et applications.

15 La figure 3 représente les positions des dix zones de mutations (Pvu II et Pst I) portées par chaque mutant du gène *ponB* utilisé pour les exemples de mises en oeuvre de l'invention.

20 La figure 4 représente la position des amorces utilisées par rapport à la séquence du gène *ponB*.

La figure 5 représente la migration des produits des réactions de RLR et de PCR des produits de ces réactions de RLR sur gel d'agarose.

25 La figure 6 représente la position des mutations par rapport aux fragments de restriction.

#### I - EXEMPLE.

30 Le procédé de l'invention a été mis en oeuvre à partir d'une banque de mutants du gène *ponB*, codant pour la PBP1b d'*E. coli* (1). Dix mutants de ce gène ont été utilisés. La séquence du gène de chaque mutant diffère de celle du gène sauvage par une zone non homologue de treize à seize bases de long constituée par la substitution de  
35 cinq codons initiaux par cinq codons alanines selon la technique décrite par Lefèvre et al. (8).

La substitution portée par chaque mutant est caractérisée par la présence d'un site unique de l'enzyme



de restriction *Pvu* II encadré par deux sites de l'enzyme *Pst* I, qui permettent de distinguer les mutants les uns des autres par leur profil de digestion avec ces endonucléases de restriction. La figure 3 représente les positions des dix zones de mutations (*Pvu* II et *Pst* I) portées par chaque mutant.

Après amplification par PCR des gènes des dix mutants, les produits PCR ont été purifiés, mélangés en quantité équimolaire pour constituer la banque. Les séquences polynucléotidiques de cette banque ont été digérées par les enzymes de restriction *Hinf* I et *Bsa* I, de façon à générer des banques de fragments de restriction. Les fragments de restriction ont alors été incubés avec la matrice sauvage, à différentes quantités, en présence d'une ligase thermostable. Après plusieurs cycles de dénaturation/hybridation/ligation, une fraction de ce mélange réactionnel a été utilisée pour réaliser une amplification PCR avec un couple d'amorces spécifiques des extrémités 5' et 3' des gènes des mutants et non spécifiques des extrémités 5' et 3' de la matrice sauvage. Le produit d'amplification a été cloné, et les clones obtenus ont été analysés pour leur profil de digestion avec l'endonucléase de restriction *Pvu* II ou *Pst* I. Les profils obtenus ont permis de déterminer quel(s) fragment(s) des mutants avai(en)t pu se recombinaison avec les autres pour reconstituer un gène entier.

## II - MATERIEL.

### 1) Souches et plasmides.

La souche MC1061 ( $F^-$  *araD*139,  $\Delta$ (*ara-leu*)7696, *galE*15, *galK*16,  $\Delta$ (*lac*)X74, *rpsL* ( $Str^R$ ), *mcrA mcrB1*, *hsdR2* ( $rk^- mk^+$ )) est dérivée d'*Escherichia coli* K12.

Le vecteur PARAPONB est issu du vecteur PARA13 (3) dans lequel le gène *ponB* portant un site de coupure thrombine (9) a été introduit entre les sites de restriction *Nco* I et *Nar* I. Le vecteur pET26b+ fait partie

de la famille des vecteurs pET développée par Studier et Moffatt (13) et commercialisés par la société NOVAGEN.

5                                    2) Oligonucléotides.

Les oligonucléotides ont été synthétisés par la société ISOPRIM (Toulouse). Les séquences des oligonucléotides sont rapportées dans le tableau I ci-dessous.

10

Tableau I

Oligo N	5' ACTGACTACCATGGCCGGAATGACCGCGAGCC 3'
Oligo E	5' CCGCGGTGGAGCGAATTCTAATTACTACCAAACATATCC 3'
Oligo M1	5' GCGCCTGAATATTGCGGAGAAAAAGC 3'
Oligo M2	5' ACAACCAGATGAAAAGAAAGGTTAATATC 3'
Oligo A1	5' ACTGACTACCATGGCC 3'
Oligo A2	5' CCGCGGTGGAGCGAATTC 3'

3) Réactifs.

15                                    Les enzymes de restriction et de modifications citées dans le tableau II ci-dessous ont été utilisées selon les recommandations des fournisseurs.

Tableau II

Enzyme	Concentration	Fournisseur
<i>Nco</i> I	10 U/ $\mu$ l	NEB
<i>Pst</i> I	20 U/ $\mu$ l	NEB
<i>Eco</i> RI	20 U/ $\mu$ l	NEB
<i>Bsa</i> I	5 U/ $\mu$ l	New England Biolabs
<i>Hinf</i> I	10 U/ $\mu$ l	New England Biolabs
<i>Pvu</i> II	10 U/ $\mu$ l	New England Biolabs
T4 DNA ligase	400 U/ $\mu$ l	New England Biolabs
Taq ADN polymérase	5 U/ $\mu$ l	PROMEGA
AMPLIGASE	100 U/ $\mu$ l	EPICENTRE

20                                    Les tampons utilisés sont rapportés dans le tableau III ci-dessous.

Tableau III

Tampons	Composition
T	Tris HCl 10 mM, pH 8.0
Polymérisation 20X	Tris HCl 100 mM pH 8,3, MgCl <sub>2</sub> 15 mM, KCl 500 mM, 1.0% Triton X100®
Restriction A 10X	500 mM NaCl, 100 mM Tris HCl pH 7,9, 100 mM MgCl <sub>2</sub> , 10 mM DTT
Restriction B 10X	1 M NaCl, 500 mM Tris HCl pH 7,9, 100 mM MgCl <sub>2</sub> , 10 mM DTT
Restriction C 10X	500 mM NaCl, 1 M Tris HCl pH 7,5, 100 mM MgCl <sub>2</sub> , 0,25% Triton X100®
AMPLIGASE 10X	200 mM Tris HCl pH 8,3, 250 mM KCl, 100 mM MgCl <sub>2</sub> , 5 mM NAD, 0,1% Triton X100®
Ligation 10X	500 mM Tris HCl pH 7,5, 100 mM MgCl <sub>2</sub> , 100 mM DTT, 10 mM ATP, 250 µg/ml BSA

### III - PREPARATION DE LA MATRICE.

10 Le gène *ponB* sauvage a été amplifié par une  
étape de réaction de PCR en utilisant comme amorces les  
oligonucléotides M1 et M2 (fig. 4). Cinq réactions de PCR  
ont été préparées en ajoutant 50 ng du plasmide pPONBPBR  
portant le gène sauvage (7) à un mélange contenant 10 µl de  
15 tampon de polymérisation, 10 µl de dNTPs 2mM, 20 pmol de  
chaque oligonucléotide M1 et M2, et 5U de Taq ADN  
polymérase, dans un volume final de 100 µl. Ces mélanges  
ont été incubés dans un thermocycleur Perkin-Elmer 9600  
selon le programme suivant : (94°C - 2 min.) - ( 94°C 15  
20 sec. - 60°C 30 sec. - 72°C 1 min.) x 29 cycles - (72°C - 3  
min.).

Le produit des cinq PCR ont été mélangés et  
déposés sur un gel d'agarose TBE 1%. Après migration et  
coloration au bromure d'éthidium du gel, la bande à 2651

pb, correspondant au produit d'amplification du gène *ponB* encadré par deux fragments de 26 pb et 90 pb respectivement, a été visualisée par trans-illumination aux ultraviolets, et découpée à l'aide d'un scalpel pour être purifiée avec le système Quiaquick (QIAGEN). La totalité de l'ADN ainsi purifié a été éluée dans 120 µl de tampon T. La concentration de cet ADN a été évaluée par un dosage spectrophotométrique à 260 nm, à 100 ng/µl.

#### IV - PREPARATION DE LA BANQUE.

##### 1) Amplification des gènes mutants.

Les gènes des dix mutants ont été amplifiés séparément par l'intermédiaire d'une réaction de PCR en utilisant les oligonucléotides N et E. Ces oligonucléotides introduisent respectivement les sites de restriction *Nco* I et *Eco* RI, permettant de cloner les produits obtenus avec ces deux sites.

Chaque réaction de PCR a été préparée en ajoutant 50 ng du plasmide portant le gène mutant à un mélange contenant 10 µl de tampon de polymérisation, 10 µl de dNTPs 2mM, 20 pmol de chaque oligonucléotide N et E, et 5U de Taq ADN polymérase, dans un volume final de 100 µl. Ce mélange a été incubé dans un thermocycleur Perkin-Elmer 9600 selon le programme suivant : (94°C - 2 min.) - ( 94°C 15 sec. - 60°C 30 sec. - 72°C 1 min.) x 29 cycles - (72°C - 3 min.).

La spécificité de l'amplification génique a été vérifiée par profil de restriction avec l'endonucléase *Pvu* II, en incubant 5 µl de chaque produit PCR 1 heure à 37°C dans un mélange contenant 3 µl de tampon de restriction A et 5U de l'enzyme *Pvu* II dans un volume final de 30 µl. 15 µl de cette réaction de digestion ont été déposés sur un gel d'agarose TBE 1%. Après migration et coloration au bromure d'éthidium, le gel a été exposé aux ultraviolets. La visualisation des fragments de restriction a permis de confirmer la

spécificité de l'amplification génique de chaque gène mutant.

En parallèle, 3 µl de chaque réaction PCR ont été déposés sur un gel d'agarose TBE 1%. Après migration, le gel a été traité comme ci-dessus. L'intensité de chaque bande a permis d'estimer que les amplifications géniques avaient eu le même rendement.

## 2) Création des banques de fragments de restriction.

50 µl de chacune des dix PCR ont été mélangés et déposés sur un gel d'agarose TBE 1%. Après migration et coloration au bromure d'éthidium, la bande à 2572 pb, correspondant au produit d'amplification des gènes des dix mutants, a été découpée à l'aide d'un scalpel et purifiée avec le système Quiaquick (QIAGEN). La totalité de l'ADN ainsi purifié a été éluée dans 120 µl de tampon T. La concentration de cet ADN a été évaluée par un dosage spectrophotométrique à 260 nm, à 100 ng/µl.

Pour générer les banques de fragments de restriction, 100 µl de cet ADN ont été incubés une heure à 50°C dans un mélange contenant 12 µl de tampon de restriction B, 1,2 µl de BSA (à 10 mg/ml), 25 U de l'enzyme Bsa I et 4 µl d'eau. Puis 2 µl du tampon de restriction B, 2 µl de BSA (à 1mg/ml), 50 U de l'enzyme Hinf I et 11,5 µl d'eau ont été rajoutés au mélange, qui a été incubé une heure à 37°C. La totalité du mélange de digestion a été purifié sur une colonne QIAquick (QIAGEN), et élué avec 30 µl de tampon T. 1 µl de cet éluat a été déposé sur gel d'agarose TBE 1% pour vérifier que la digestion avait été totale, et qu'elle avait généré 6 fragments de restriction, et par conséquent six banques de fragments, de 590 pb, 500 pb, 472 pb, 438 pb, 298 pb et 274 pb. La concentration de cet ADN a été évaluée (par un dosage spectrophotométrique à 260 nm) à 250 ng/µl.

## V - RLR (Recombinating Ligation Reaction).

La réaction de RLR (Recombining Ligation Reaction) a été réalisée en incubant des quantités déterminées de fragments de restriction *Hinf* I - *Bsa* I des gènes des dix mutants avec la matrice complète (ie le gène *ponB* sauvage), en présence d'une ADN ligase thermostable. Le tableau IV ci-dessous rapporte la composition des mélanges de RLR.

Tableau IV

	RLR 1	RLR 2	RLR3	RLR4	T-
Fragments <i>Hinf</i> I - <i>Bsa</i> I des dix mutants (100 ng/ $\mu$ l)	0,5 $\mu$ l	1 $\mu$ l	2 $\mu$ l	5 $\mu$ l	5 $\mu$ l
Matrice <i>ponB</i> sauvage (100 ng/ $\mu$ l)	0,6 $\mu$ l	1,2 $\mu$ l	2,4 $\mu$ l	6 $\mu$ l	6 $\mu$ l
Tampon AMPLIGASE 10X	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l
AMPLIGASE (25 U/ $\mu$ l)	1 $\mu$ l	1 $\mu$ l	1 $\mu$ l	1 $\mu$ l	-
H <sub>2</sub> O	qsp 20 $\mu$ l	qsp 20 $\mu$ l	qsp 20 $\mu$ l	qsp 20 $\mu$ l	qsp 20 $\mu$ l

Le témoin négatif est identique à la réaction de RLR4, mais ne contient pas d'ADN ligase thermostable. Ces différents mélanges ont été recouverts d'une goutte d'huile minérale et incubés dans un thermocycleur Perkin-Elmer 9600 dans des microtubes de 200  $\mu$ l selon le programme suivant : (94°C 5 min.) - (94°C 1 min. - 65°C 4 min.) x 35 cycles.

10  $\mu$ l de chaque réaction de RLR ont alors été ajoutés à un mélange de réaction PCR contenant 10  $\mu$ l de tampon de polymérisation, 10  $\mu$ l de dNTPs 2 mM, 40 pmol de chaque oligonucléotide A1 et A2, et 5 U de Taq ADN polymérase dans un volume final de 100  $\mu$ l. Ce mélange a été incubé dans un thermocycleur Perkin-Elmer 9600 selon le programme suivant : (94°C 5 min.) - (94°C 30 sec. - 46°C 30 sec. - 72°C 1 min.) x 29 cycles - (72°C 2 min.). Cette réaction de PCR a permis d'amplifier spécifiquement les

produits de ligation formés au cours de la réaction de RLR, sans amplifier la matrice, puisque les oligonucléotides A1 et A2 ne peuvent pas s'hybrider sur celle-ci, comme montré sur la figure 4.

5                    5 µl de chaque réaction de RLR et 10 µl de chacune des réactions de PCR précédentes ont été déposés sur un gel d'agarose TBE 1%. Après coloration au bromure d'éthidium, le gel a été exposé aux ultraviolets, comme montré à la figure 5.

10                   L'analyse de ce gel révèle que seule la réaction de RLR4 contient, comme le témoin négatif, des fragments de restrictions encore visibles (pistes 4 et 5).

15                   L'absence de produit PCR pour le témoin négatif (piste 10) révèle non seulement que la réaction PCR est spécifique (pas d'amplification de la matrice complète), mais aussi que les fragments de restriction présents dans le mélange ne peuvent pas se substituer aux amorces pour générer un produit PCR contaminant dans les conditions choisies. En parallèle, la présence d'une bande unique à environ 2500 pb sur les pistes 6, 7 et 8 démontre qu'un produit de RLR a pu être amplifié par PCR pour les réactions de RLR1, 2 et 3. Ces trois réactions de RLR ont donc permis de reconstituer un ou des gènes complets à partir de six banques de fragments de restrictions.

## 25                   VI - ANALYSE DES PRODUITS D'AMPLIFICATION DES REACTIONS DE RLR.

### 30                   1) Clonage.

35                   Les produits d'amplification PCR des réactions de RLR 1, 2 et 3 ont été purifiés avec le système Wizard PCR Preps (PROMEGA) et élués dans 45 µl de tampon T. 6 µl de chaque PCR purifiée ont été incubés 1 heure à 37°C dans un mélange contenant 3 µl de tampon de restriction C, 3 µl de BSA (à 1mg/ml), 20 U de l'enzyme Eco RI, 10 U de l'enzyme Nco I et 15 µl d'eau.

En parallèle, deux vecteurs (pARAPONB et pET26b+) ont été préparés pour le clonage. Ces vecteurs ont

été linéarisés en incubant 3 µg de ces plasmides 2 heures à 37°C, dans un mélange contenant 3 µl de tampon de restriction C, 3 µl de BSA (à 1mg/ml), 20 U de l'enzyme *Eco* RI, 10 U de l'enzyme *Nco* I et 19 µl d'eau.

Les vecteurs linéarisés ainsi que les PCR digérées ont été purifiés sur gel d'agarose TBE 1% avec le système QIAquick (QUIAGEN). Chaque vecteur ou chaque PCR digérée a été éluée dans 30 µl de tampon T.

La ligation de chaque PCR digérée avec l'un ou l'autre des vecteurs a été réalisée selon les conditions décrites dans le tableau V ci-dessous, et incubée à 16°C pendant 16 heures.

Tableau V

	Ligation avec le vecteur pARAPONB				Ligation avec le vecteur pET26b+			
	LpAR1	LpAR2	LpAR3	TLpAR	LpET1	LpET2	LpET3	TLpET
PCR amplification RLR 1 digérée <i>Nco</i> I - <i>Eco</i> RI	4 µl	-	-	-	4 µl	-	-	-
PCR amplification RLR 2 digérée <i>Nco</i> I - <i>Eco</i> RI	-	4 µl	-	-	-	4 µl	-	-
PCR amplification RLR 3 digérée <i>Nco</i> I - <i>Eco</i> RI	-	-	4 µl	-	-	-	4 µl	-
Vecteur pARAPONB digérée <i>Nco</i> I - <i>Eco</i> RI	1 µl	1 µl	1 µl	1 µl	-	-	-	-
Vecteur pET26b+ digérée <i>Nco</i> I - <i>Eco</i> RI	-	-	-	-	1 µl	1 µl	1 µl	1 µl
Tampon de ligation	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl
Ligase	1 µl	1 µl	1 µl	1 µl	1 µl	1 µl	1 µl	1 µl
H <sub>2</sub> O	12 µl	12 µl	12 µl	16 µl	12 µl	12 µl	12 µl	16 µl

200 µl de cellules MC1061 chimiocompétentes (4) ont été transformées avec 10 µl de chaque ligation par un



choc thermique (5), et les cellules ainsi transformées ont été étalés sur un milieu de sélection.

Aucun clone n'a été obtenu après transformation des témoins de ligation TL pAR et TL pET, indiquant ainsi que les vecteurs PARAPONB et pET26b+ linéarisés *Nco* I - *Eco* RI ne peuvent pas subir une ligation intramoléculaire.

## 2) Criblage par PCR.

Un premier criblage des clones obtenus après transformation des ligations avec le vecteur PARAPONB a été réalisée par PCR. 42 colonies, 14 de chaque ligation LpAR1, LpAR2 et LpAR3, ont été resuspendues individuellement dans un mélange de PCR contenant 5 µl de tampon de polymérisation, 40 pmol de chaque oligonucléotide A1 et A2, 5 µl de dNTPs 2mM et 5U de Taq ADN polymérase dans un volume final de 50 µl. Un témoin négatif a été constitué en ajoutant au mélange de PCR 50 ng du plasmide pBR322 à la place d'une colonie. Ces 43 tubes ont été incubés dans un thermocycleur Perkin-Elmer 9600 selon le programme suivant : (94°C 5 min.) - (94°C 30 sec. - 46°C 30 sec. - 72°C 1 min.) x 29 cycles - (72°C 2 min.). 5 µl de chacune de ces réactions PCR ont ensuite été incubés 1 heure à 37°C dans un mélange contenant 2 µl de tampon de restriction A, 2 µl de BSA (à 1 mg/ml) et 5 U de l'enzyme de restriction *Pvu* II dans un volume final de 20 µl.

10 µl de chacune de ces digestions ont été déposées sur un gel d'agarose TBE 1% en parallèle de 5 µl de chaque PCR non digérée (ce qui permet de ne pas confondre d'éventuelles bandes non spécifiques de la PCR avec un fragment obtenu par digestion de restriction). Après migration et coloration au bromure d'éthidium de ce gel, les bandes issues de la digestion par l'enzyme *Pvu* II ont été analysées afin de déterminer quel(s) fragment(s) des mutants initiaux s'étaient associé(s) avec les autres pour reconstituer un gène entier. Ce criblage révèle la présence de 27 gènes portant une mutation, 7 gènes portant deux mutations et de 8 gènes ne portant plus de mutation.

3) Criblage par minipréparation d'ADN plasmidique.

Le deuxième criblage a été effectué en réalisant une extraction de l'ADN plasmidique (5) de 21 clones issus de la transformation des ligations avec le vecteur pET26b+ (7 clones de chaque ligation). 5 µl de l'ADN plasmidique ainsi obtenu pour chaque clone a été incubé 1 heure à 37°C en présence d'un mélange contenant 1 µl de tampon de restriction C, 6 U de l'enzyme *Pst* I, 3 U de l'enzyme *Nco* I et 6 U de l'enzyme *Eco* RI dans un volume final de 10 µl. 5 µl de chacune de ces digestions ont été déposées sur un gel d'agarose TBE 1%. Après migration et coloration au bromure d'éthidium de ce gel, les bandes issues de la digestion par l'enzyme *Pst* I ont été analysées afin de déterminer quel(s) fragment(s) des mutants initiaux s'étaient associé(s) avec les autres pour reconstituer un gène entier. Ce criblage révèle la présence de 13 gènes portant une mutation, 5 gènes portant deux mutations et de 3 gènes ne portant plus de mutation.

4) Analyse statistique des recombinaisons.

En fonction de la position de chaque mutation par rapport aux sites de coupure des enzymes *Hinf* I et *Bsa* I, comme représenté à la figure 6, il est possible de calculer la probabilité d'obtenir au cours de la réaction de RLR la création d'un gène portant 0, 1, 2, 3 ou 4 des mutations des gènes initiaux.

Ainsi, en considérant que la réaction de RLR est totalement aléatoire les probabilités P sont les suivantes :

$$P(0 \text{ mutation}) = \prod_{i=6}^9 \left( \frac{i}{10} \right) = 30,24\%$$

$$P(1 \text{ mutation}) = \sum_{n=1}^4 \left[ \frac{n}{10-n} \prod_{i=1}^4 \left( \frac{10-i}{10} \right) \right] = 44,04\%$$

$$P(2 \text{ mutations}) = \sum_{n=1}^4 \left[ \sum_{a=1}^{4-n} \left( \frac{10-a}{a} \right) \left( \frac{10-(a+n)}{a+n} \right) \prod_{i=1}^4 \left( \frac{i}{10} \right) \right] = 21,44\%$$

$$P(3 \text{ mutations}) = \sum_{n=1}^4 \left[ \left( \frac{10-n}{n} \right) \prod_{i=1}^4 \left( \frac{i}{10} \right) \right] = 4,04\%$$

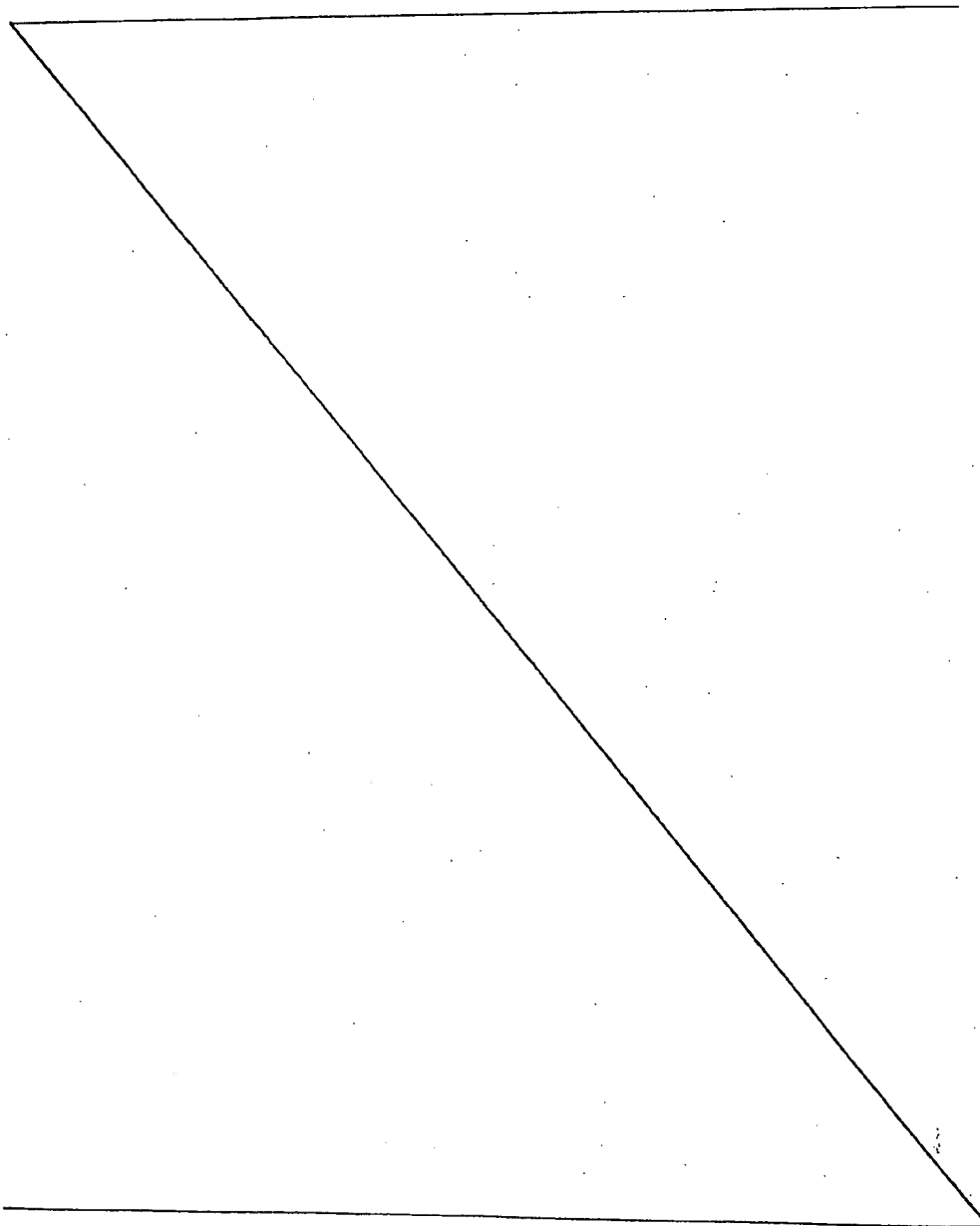
$$P(4 \text{ mutations}) = \prod_{i=1}^4 \left( \frac{i}{10} \right) = 0,24\%$$

Les deux criblages effectués donnent des résultats proches de ces prévisions statistiques, comme rapporté dans le tableau VI ci-dessous, indiquant ainsi que la réaction de RLR est quasi aléatoire. Une proportion légèrement plus élevée de gènes portant une mutation, au détriment des gènes portant zéro mutation est observée. Ce phénomène pourrait être attribué à une faible toxicité du gène *ponB* déjà observée et à la légère fuite des vecteurs d'expression *PARAPONB* et *pET26b+*, qui favoriseraient la sélection de gènes portant une mutation inactivante.

Tableau VI

%	0 mutation	1 mutation	2 mutations	3 mutations	4 mutations
Statistique	30,24	44,04	21,44	4,04	0,24
Criblage	21	63	16	0	0
PCR					

Criblage mini- préparation	14	62	24	0	0
----------------------------------	----	----	----	---	---



RÉFÉRENCES BIBLIOGRAPHIQUES

5 1) Broome-Smith J.K., Edelman A., Yousif S. et  
Spratt B.G., (1985), The nucleotide sequence of the *ponA*  
and *ponB* genes encoding penicillin-binding proteins 1A and  
1B of *Escherichia coli* K12, Eur. J. Biochem., 147, 437-446.

10 2) Cadwell R.C. et Joyce G., 1992,  
Randomization of genes by PCR mutagenesis, PCR Methods and  
Application, 2, 28-33.

15 3) Cagnon C., Valaverde V. et Masson J.-M.,  
(1991), A new family of sugar inducible expression vectors  
for *Escherichia coli*, Prot. Eng., 4, 843-847.

20 4) Hanahan D., (1985), Techniques for  
transformation of *Escherichia coli*, dans DNA cloning : a  
practical approach, Glover D.M. (Ed), IRL Press, Oxford vol  
I, 109-135.

5) Maniatis T., Fritsch E.F. et Sambrook J.,  
(1982), Molecular cloning. A laboratory Manual, Cold Spring  
Harbor Laboratory, Cold Spring Harbor, NY.

25 6) Landt et al., Gene, 96, 125-128, 1990

30 7) Lefèvre F., Analyse topologique de la  
Penicillin Binding Protein 1b d'*Escherichia coli*, 1997,  
Thèse.

8) Lefèvre F., Rémy M.H. et Masson J.M., 1997  
(a), Alanine-stretch scanning mutagenesis : a simple and  
efficient method to probe protein structure and function,  
Nuc. Acids Res., 25, 447-448.

35 9) Lefèvre F., Rémy M.H. et Masson J.M., 1997  
(b), Topographical and functional investigation of  
*Escherichia coli* Penicillin-Binding Protein 1b by alanine

stretch scanning mutagenesis, J. Bacteriol., 179, 4761-4767.

5 10) Lyamichev V., Mast A.L., Prudent J.R., Kaiser M.W., Takova T., Kwiatkowski R.W., Sander T.J., de Arruda M., Arco D.A., Neri B.P. et Brown M.A.D., 1999, Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes, Nature Biotechnology, 17, 292-296.

10 11) Picard et al., Nuc. Acids Res., 22, 2587-2591, 1994.

12) Stemmer W.P.C., (1994), Rapid evolution of a protein in vitro by DNA shuffling, Nature, 370, 141-144.

15 13) Studier F.W. et Moffatt B.A., 1986, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J. Mol. Biol, 189, 113-130.

20 14) Zhao H., Giver L., Shao Z., Affholter J.A. et Arnold F., 1998, Molecular evolution by staggered extension process (StEP) in vitro recombination, Nature Biotech., 16, 258-261.

## REVENDEICATIONS

1) Procédé d'obtention *in vitro* de séquences polynucléotidiques recombinaées à partir d'une banque de séquences polynucléotidiques, caractérisé en ce qu'il comprend les étapes suivantes :

a) la fragmentation d'une banque de séquences polynucléotidiques double-brins,

b) la dénaturation des fragments éventuellement en présence d'une ou plusieurs matrice(s) d'assemblage,

c) l'hybridation desdits fragments avec une ou plusieurs matrice(s) d'assemblage si celle(s)-ci n'est (ne sont) pas présente(s) à l'étape (b),

d) la ligation desdits fragments pour obtenir des séquences polynucléotidiques recombinaées,

e) la sélection des séquences polynucléotidiques recombinaées présentant des propriétés avantageuses par rapport aux propriétés correspondantes à une ou plusieurs séquences de référence.

2) Procédé selon la revendication 1, caractérisé en ce qu'il comprend à l'issue de l'étape (d) et avant l'étape (e), la répétition des étapes (b), (c) et (d) avec les fragments ligués et non ligués issus de l'étape (d).

3) Procédé selon l'une des revendications 1 ou 2, caractérisé en ce qu'il comprend la séparation des séquences polynucléotidiques recombinaées de la ou des matrice(s) d'assemblage avant l'étape (e).

4) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comprend l'amplification des séquences polynucléotidiques recombinaées double brin avant l'étape (e).

5) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il

comprend avant l'étape (e), le clonage de séquences polynucléotidiques recombinées éventuellement après séparation des brins recombinés de la ou des matrices et obtention du double brin correspondant.

5

6) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que les extrémités des fragments générés à l'étape (a) sont telles qu'il puisse y avoir hybridation adjacente de ces extrémités sur la ou les matrice(s) d'assemblage à l'étape (c) et ligation de ces fragments les uns avec les autres à l'étape (d).

10

7) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que les séquences polynucléotidiques de la banque initiale présentent des zones d'homologie soit entre elles, soit avec les matrices d'assemblage, de façon à générer à l'étape (a) des extrémités de fragments qui permettent l'hybridation adjacente de ces extrémités sur la ou les matrice(s) d'assemblage à l'étape (c) et ligation de ces fragments les uns avec les autres à l'étape (d).

15

20

8) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que les étapes (c) et (d) sont réalisées simultanément.

25

9) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que la fragmentation des séquences polynucléotidiques à l'étape (a) se fait de manière contrôlée ou de manière aléatoire.

30

10) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'étape (a) consiste à soumettre les séquences polynucléotidiques de la banque initiale à une hydrolyse par l'action d'une ou plusieurs enzymes de restriction.

35



11) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que le degré de recombinaison et la position des points de recombinaison des séquences polynucléotidiques recombinaison sont déterminés par la fragmentation de l'étape (a) si elle est réalisée de manière contrôlée.

12) Procédé selon l'une quelconque des revendications 1 à 9 caractérisé en ce que la fragmentation aléatoire des séquences polynucléotidiques à l'étape (a) se fait par tout moyen enzymatique ou mécanique connu.

13) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'on ajoute à l'étape (c) et/ou à l'étape (d) des enzymes capables de reconnaître et de couper de manière spécifique les extrémités non hybridées des fragments, lorsque lesdites extrémités recouvrent d'autres fragments hybridés sur la même matrice..

14) Procédé selon la revendication 13, caractérisé en ce que l'on ajoute à l'étape (c) et/ou à l'étape (d) l'enzyme Flap endonucléase.

15) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'on utilise à l'étape (d) une ligase active à haute température et de préférence thermostable.

16) Procédé selon les revendications 13 et 14, caractérisé en ce que les endonucléases capables de reconnaître et de couper de manière spécifique les extrémités non hybridées des fragments ajoutées à l'étape (c) et/ou à l'étape (d) ont les mêmes propriétés de thermorésistance et d'activité à haute température que la ligase utilisée à l'étape (d).

17) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que la banque initiale de séquences polynucléotidiques est générée à partir d'un gène sauvage par étapes de mutagénèse dirigée successive, par PCR error prone, par mutagénèse aléatoire chimique, par mutagénèse aléatoire *in vivo*, ou en combinant des gènes de familles proches ou distinctes au sein de la même espèce ou d'espèces différentes de façon à disposer d'une variété de séquences polynucléotidiques dans la banque initiale.

18) Procédé selon l'une quelconque des revendications 1 à 16, caractérisé en ce que la banque initiale de séquences polynucléotidiques double-brins est constituée de séquences synthétiques qui seront fragmentées à l'étape (a) ou qui peuvent constituer les fragments de l'étape (a)

19) Procédé selon l'une quelconque des revendications 1 à 11, 13 à 18, caractérisé en ce que l'étape (a) consiste à soumettre la banque initiale à une hydrolyse par l'action d'enzymes de restriction ayant un grand nombre de sites de coupure sur les séquences polynucléotidiques de la banque initiale, ou en combinant plusieurs enzymes de restriction.

20) Procédé selon l'une quelconque des revendications 1 à 9 et 12 à 18, caractérisé en ce que l'étape (a) consiste en un traitement aléatoire avec la DNase I d'une banque initiale de séquences polynucléotidiques double brins partiellement hétérologues.

21) Procédé selon l'une quelconque des revendications 12 ou 20, caractérisé en ce que l'on utilise des fragments générés par un traitement de façon aléatoire comme matrices les uns pour les autres, pour

l'hybridation au cours de l'étape (c) ou de la réaction de RLR des étapes (c) et (d) simultanées.

22) Procédé selon l'une des revendications 10 ou 19, caractérisé en ce que l'étape (b) est réalisée en combinant au moins deux banques de fragments distinctes générées séparément à l'étape (a) à partir de la même banque initiale par un traitement avec des enzymes de restriction différentes.

23) Procédé selon la revendication 22, caractérisé en ce que l'on utilise les fragments obtenus à l'étape (a) par un traitement avec des enzymes de restriction comme matrices les uns pour les autres, pour l'hybridation au cours de l'étape (c) ou de la réaction de RLR des étapes (c) et (d) simultanées.

24) Procédé selon l'une quelconque des revendications 1 à 9, 11, 13 à 17 caractérisé en ce que les fragments de l'étape (a) sont obtenus par des réactions d'amplification menées sur les séquences polynucléotidiques de la banque initiale.

25) Procédé selon la revendication 24, caractérisé en ce que les réactions d'amplification sont réalisées avec des oligonucléotides amorces permettant de générer des fragments dont les extrémités sont adjacentes tout au long de la séquence d'assemblage.

26) Procédé selon la revendication 24, caractérisé en ce que les réactions d'amplification sont réalisées avec des oligonucléotides amorces permettant de générer des fragments ayant des séquences communes, lesdits fragments servant de matrice d'assemblage les uns pour les autres à l'étape (b) ou à l'étape (c).

27) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que à l'étape

(a) la banque initiale est fragmentée en n fragments, n étant supérieur ou égal à trois.

5                   28) Procédé selon l'une quelconque des revendications 1 à 18, 24 à 27, caractérisé en ce que la matrice d'assemblage à l'étape (b) ou (c) est une séquence polynucléotidique issue de la banque initiale ou une séquence consensus de ladite banque, simple ou double brin.

10                   29) Procédé selon l'une quelconque des revendications 1 à 19, 24 à 27, caractérisé en ce que l'on utilise comme matrice d'assemblage aux étapes (b) ou (c) des oligonucléotides simple ou double brin juste complémentaires de l'extrémité 3' d'un fragment et 5' du  
15 fragment adjacent.

                  31) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'on ajoute à l'étape (b) ou (c) en plus de la matrice, des  
20 oligonucléotides, simple ou double brin, de longueur variable.

                  32) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que, avant  
25 l'étape (e), on sépare les séquences polynucléotidiques recombinaées de la matrice d'assemblage grâce à un marqueur présent sur la matrice d'assemblage ou sur les séquences polynucléotidiques recombinaées.

30                   33) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que les séquences polynucléotides recombinaées obtenues à l'étape (d) et éventuellement clonées sont criblées par tout moyen approprié pour sélectionner les séquences  
35 polynucléotidiques recombinaées ou les clones présentant des propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence.

34) Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que la banque de séquences polynucléotidiques initiale est constituée par une ou plusieurs banques restreintes préparées par un procédé selon l'une quelconque des revendications 1 à 33, éventuellement mélangées avec d'autres séquences polynucléotidiques.

35) Une séquence polynucléotidique recombinée présentant une ou plusieurs propriétés avantageuses par rapport aux propriétés correspondantes de séquences de référence, obtenue et sélectionnée par un procédé selon l'une quelconque des revendications 1 à 34.

36) Un vecteur contenant une séquence polynucléotidique selon la revendication 35.

37) Un hôte cellulaire transformé par une séquence polynucléotidique recombinée selon la revendication 35 ou par un vecteur selon la revendication 36.

38) Une protéine codée par une séquence polynucléotidique recombinée selon la revendication 37.

39) Une banque constituée de séquences polynucléotidiques recombinées selon la revendication 35, ou de vecteur selon la revendication 36, ou d'hôtes cellulaires selon la revendication 37, ou de protéines selon la revendication 38.

Figure 1 A

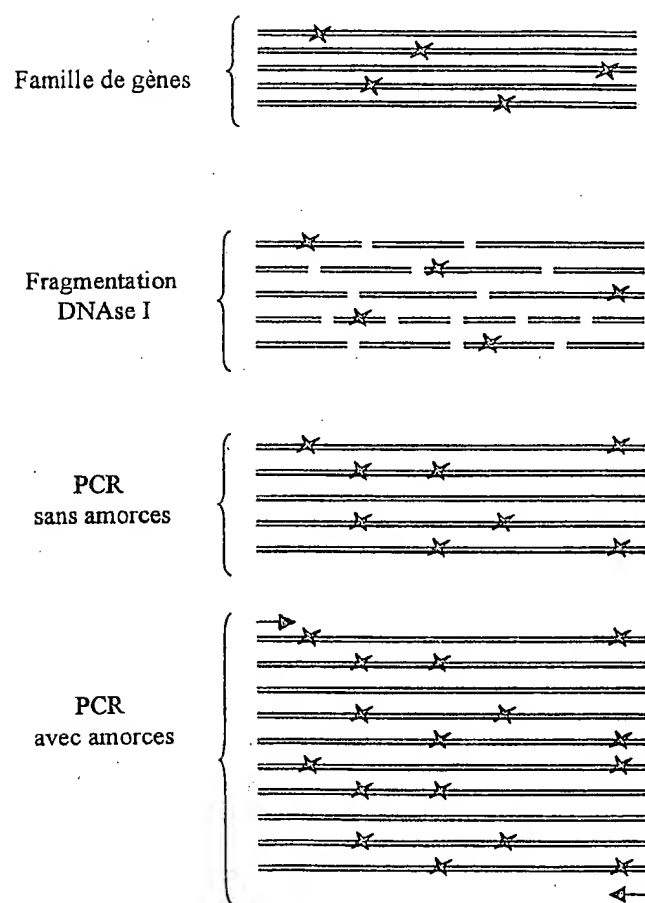
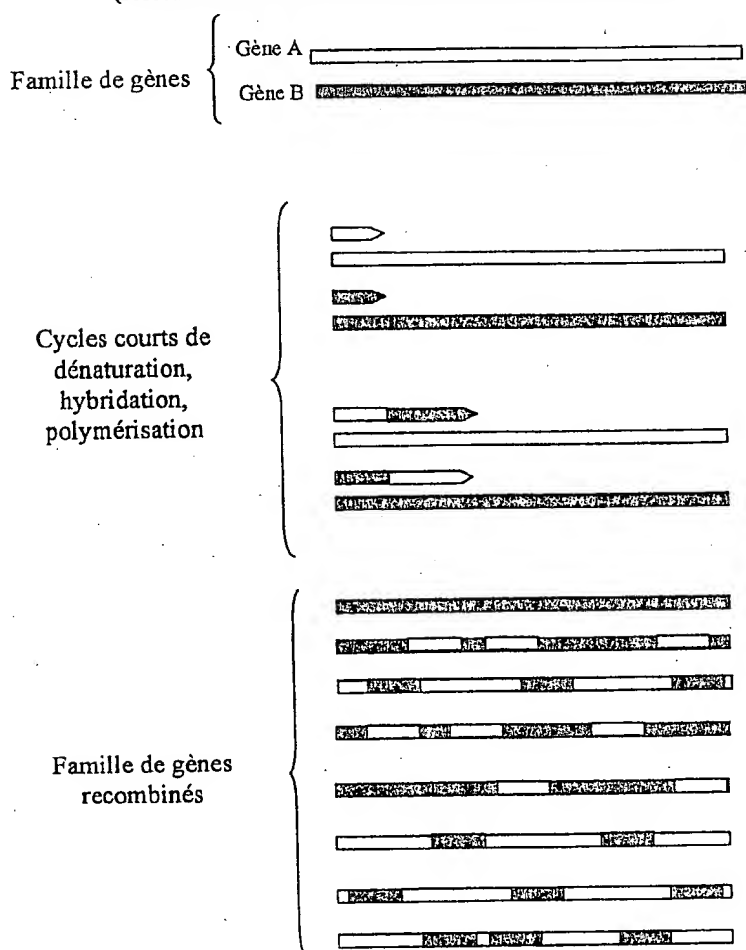


Figure 1 B

(Procédé réalisé en double brin, mais illustré ici sur un seul brin)



3/9

Fig.2

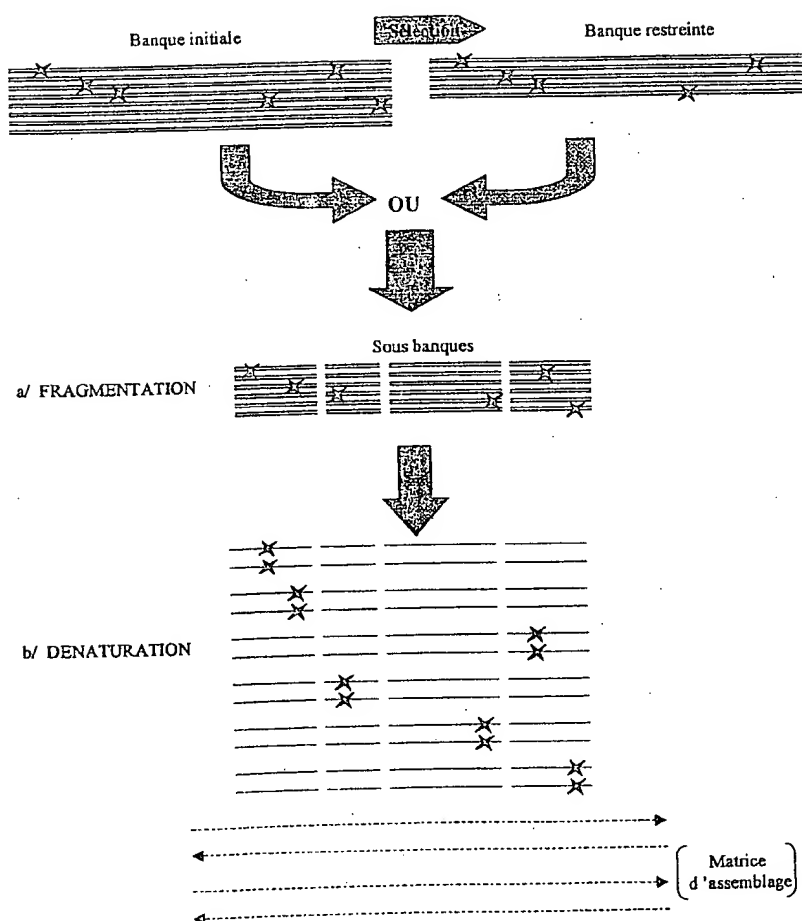
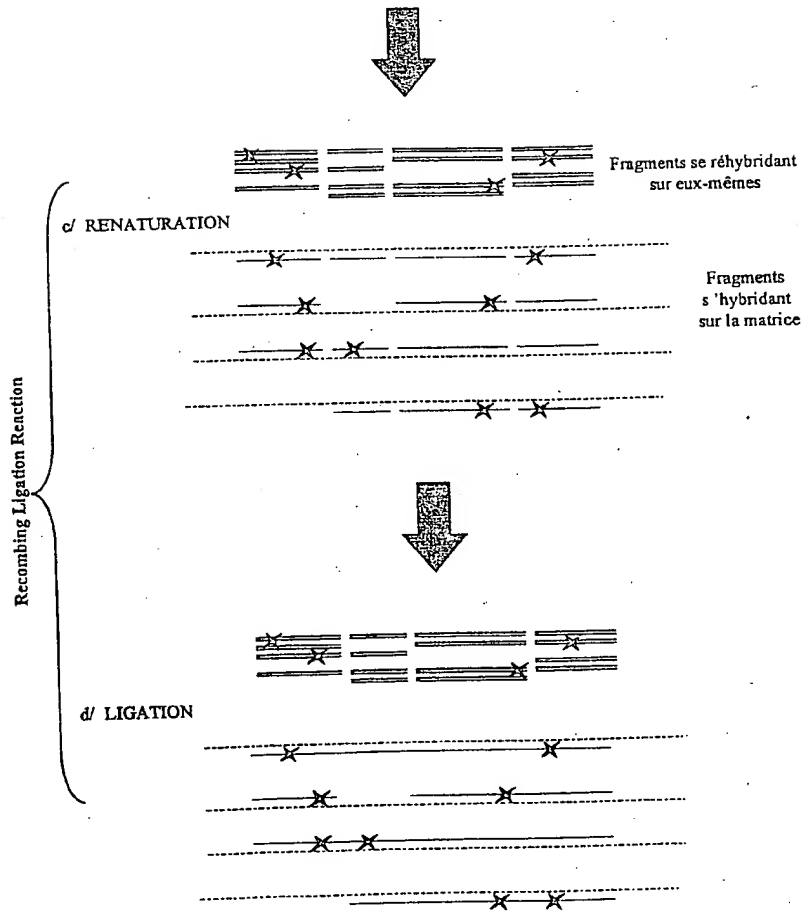


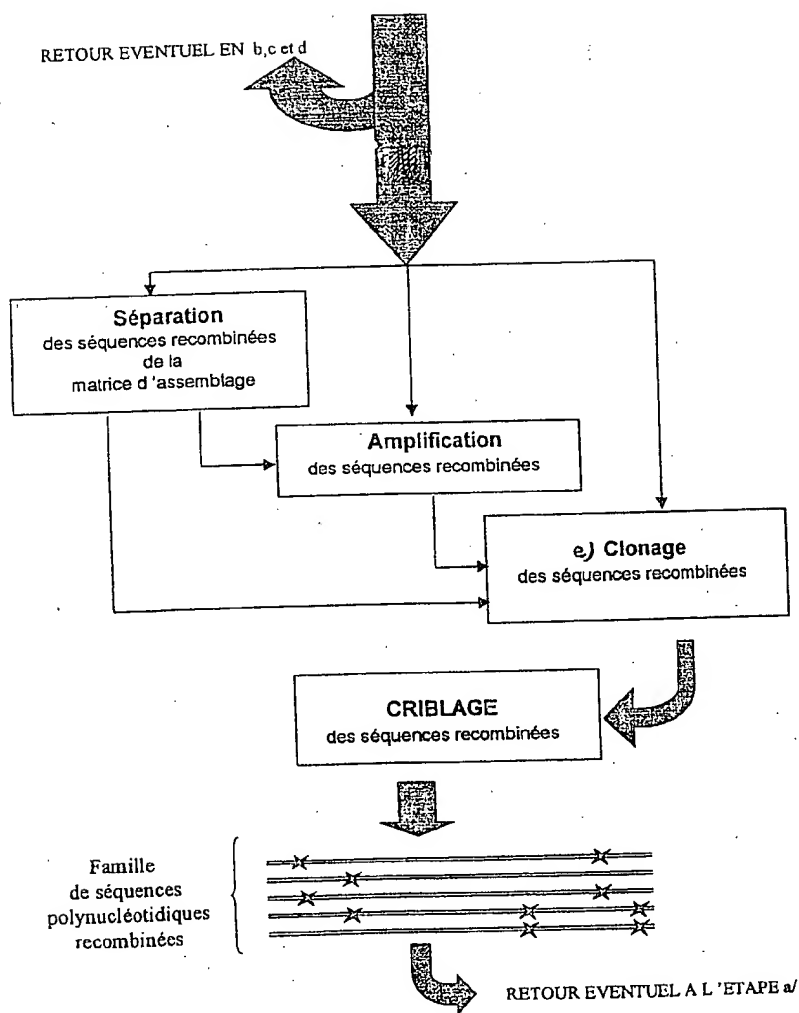


Fig.2 Suite 1



5/9

Fig.2 Suite2



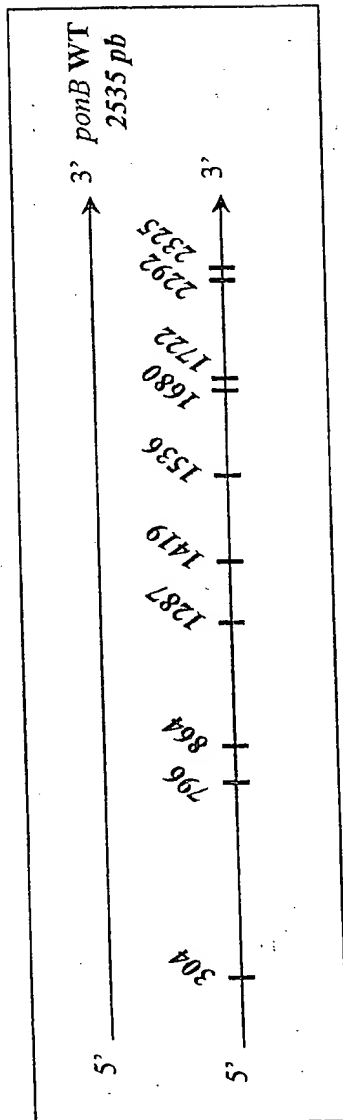


Figure 3: Position des dix zones de mutations (sites *Pvu* II et *Pst* I)

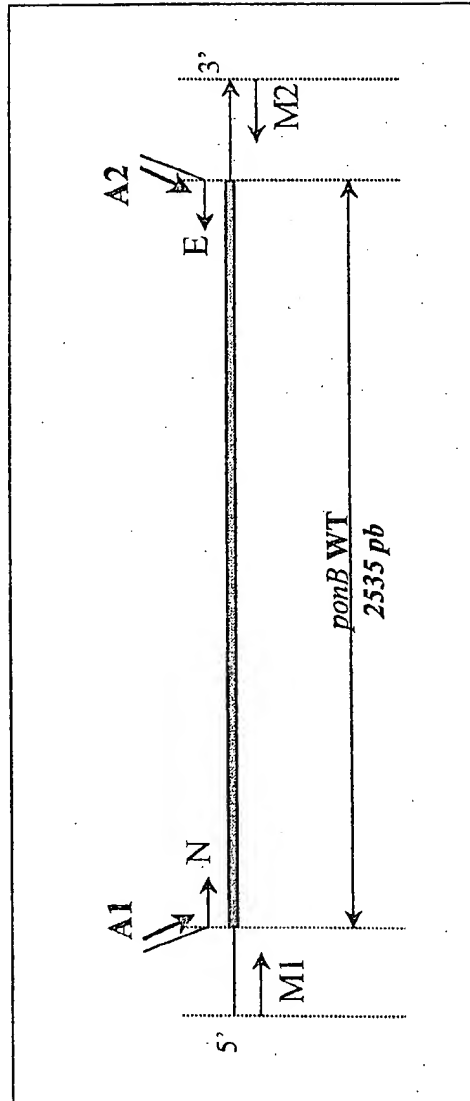


Figure 4 : Position des amorces utilisées par rapport à la séquence du gène *ponB*

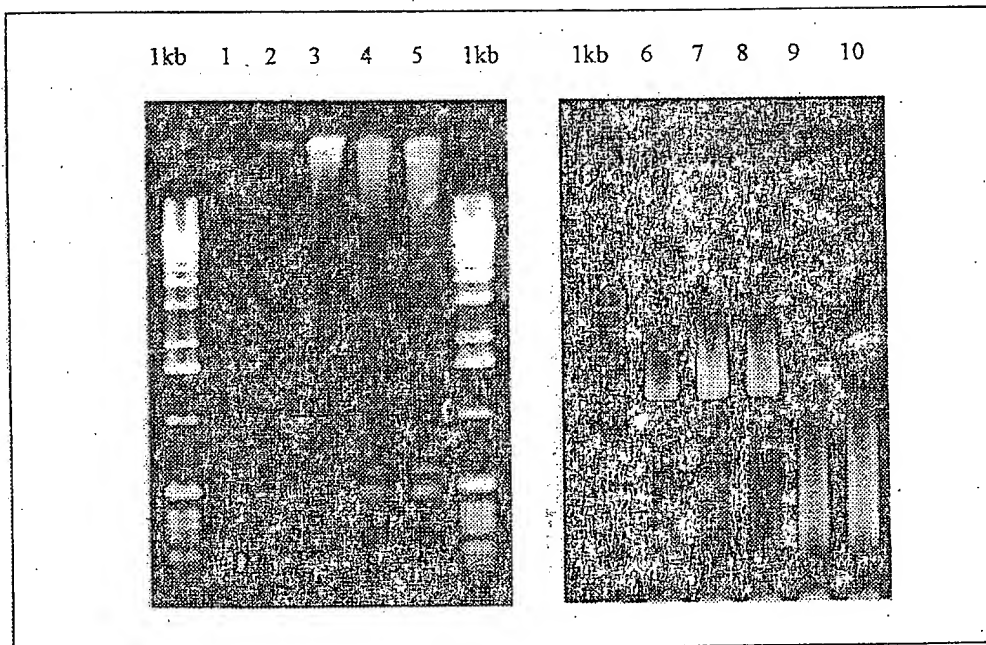


Fig. 5 : Migration des réactions de RLR et des amplifications PCR de ces réactions

*Pistes :*

1/ RLR 1	6/ PCR RLR 1
2/ RLR 2	7/ PCR RLR 2
3/ RLR 3	8/ PCR RLR 3
4/ RLR 4	9/ PCR RLR 4
5/ Témoin RLR	10/ PCR Témoin RLR

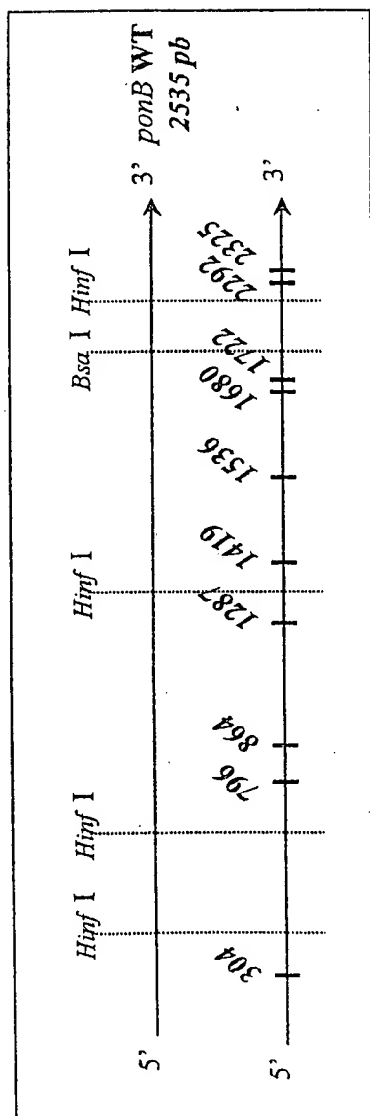


Figure 6 : Position des mutations par rapport aux fragments de restriction

PCT/FR 99/01973

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/10 C12N15/31 C12N1/21 C07K14/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STEMMER W P C: "DNA SHUFFLING BY RANDOM FRAGMENTATION AND REASSEMBLY: IN VITRO RECOMBINATION FOR MOLECULAR EVOLUTION" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 91, 1 October 1994 (1994-10-01), pages 10747-10751, XP002912215 the whole document	1-39
Y	WO 98 27230 A (MAXYGEN INC ; PATTEN PHILLIP A (US); STEMMER WILLEM P C (US)) 25 June 1998 (1998-06-25) page 13, line 15 -page 15, line 21	1-39
Y	WO 98 15567 A (KONCZ CSABA ; SCHELL JEFF (DE); STRIZHOV NICOLAI (DE); VITALITY BIO) 16 April 1998 (1998-04-16) page 2, last paragraph -page 4	1-39
-/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "S" document member of the same patent family		
Date of the actual completion of the international search  5 November 1999		Date of mailing of the international search report  22/11/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  Van der Schaal, C

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 99/01973

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WEISBERG E P ET AL: "SIMULTANEOUS MUTAGENESIS OF MULTIPLE SITES: APPLICATION OF THE LIGASE CHAIN REACTION USING PCR PRODUCTS INSTEAD OF OLIGONUCLEOTIDES" BIOTECHNIQUES, vol. 15, no. 1, 1 July 1993 (1993-07-01), pages 68-70, 72 - 74,, XP000385832 the whole document ----	1-39
Y	WO 98 32845 A (CRIPPS JOANNA E ;SOEDERLIND ULF HANS ESKIL (SE); BIOINVENT INT AB) 30 July 1998 (1998-07-30) the whole document ----	1-39
P,Y	LYAMICHEV V ET AL: "Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes." NATURE BIOTECHNOLOGY, (1999 MAR) 17 (3) 292-6. , XP002121762 cited in the application abstract; figure 1 ----	14,16
P,Y	PATENT ABSTRACTS OF JAPAN vol. 1999, no. 08, 30 June 1999 (1999-06-30) & JP 11 075849 A (AGENCY OF IND SCIENCE &AMP;TECHNOL), 23 March 1999 (1999-03-23) abstract ----	14,16
A	ZHAO H ET AL: "Molecular evolution by staggerd extension process (StEP) in vitro recombination" NATURE BIOTECHNOLOGY, vol. 16; March 1998 (1998-03), pages 258-261, XP002104755 cited in the application the whole document -----	

1